

Holiday Travel Analysis

Hanna Mortenson
Faculty Mentor: Dr. Robert Krueger

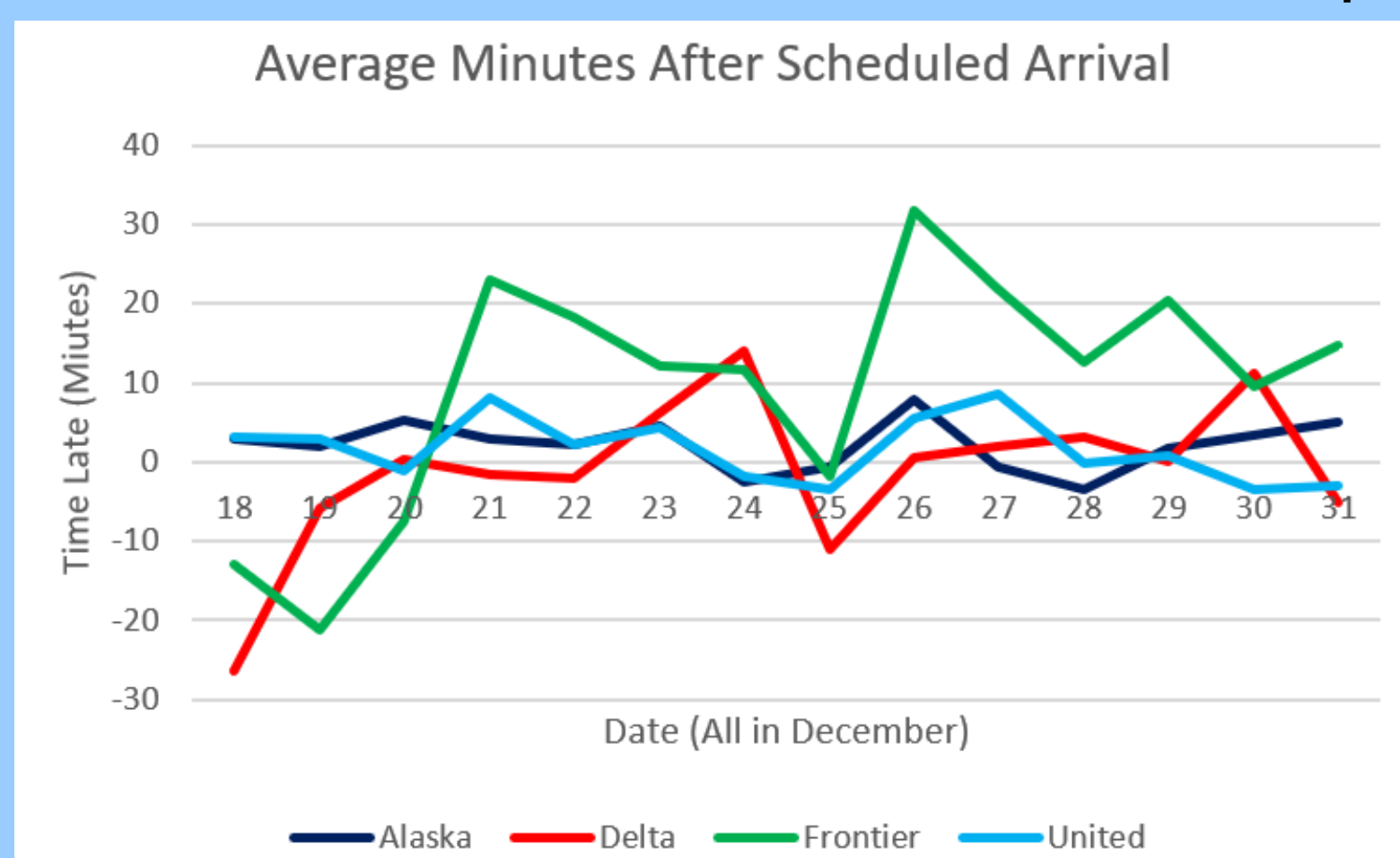


Introduction: Being stuck at the airport is never a fun thing, especially around the holidays when you're just trying to make it home to be with the family. Is there any way to avoid travel delays? On December 18th, 2019 TSA estimated that between December 19th-January 5th, "42 million passengers will travel through security screening checkpoints nationwide". Mass number of flyers, weather, etc. all play different rolls in travel delays. This study originally set out to determine which day between December 18th and December 31st would there be least amount of travel delays expected.

Airports were chosen from CNN Travels Top 5 Busiest Airports in the US, which are Hartsfield-Jackson Atlanta, Los Angeles, Chicago O'Hare, Dallas/Fort Worth, and Denver. Airlines were chosen from Forbes 2018 Top 5 Airlines, which are Alaska, Frontier, Delta, United and Southwest. However, Southwest's data was eliminated since it did not fly to all 5 airports. The data that was collected was the scheduled and actual arrival time of all flights into the 5 airports from December 18th-December 31st in the years 2014-2018.

Research Methods: The first step in this research process was mentioned above in the Introduction, the data had to be narrowed down to an amount that was workable. Next was cleaning and sorting the data into content that could be used for analysis. This process took the longest. Original data from the Bureau of Transportation Statistics came as data for one airport and one airline at a time. The raw data given was scheduled arrival time and actual arrival time, which was then transformed into average minutes that the flight arrived after the scheduled arrival time. Five different Excel formulas were needed for this process to occur to eliminate all errors.

From there the data was put altogether in a pivot table to easily be able to sort between different categories such as date or airline. Next a basic level of analysis started such as the graph pictured here. This shows the average minutes after scheduled arrival time by airline by day. Just with a brief glance, it seems as though one of these airlines does not follow a similar pattern as the others.



A Two-Way Analysis of Variance (ANOVA) was used to test if the delay averages were equal between airlines or equal between days. ANOVA is a statistical test which is used to see if the means of groups are significantly statistically different from each other. Further statistical analysis followed with Paired T-Tests. Paired T-Tests are used to determine whether the mean difference between two groups are less than zero, equal zero, or are greater than zero, depending on your choice of claim.

This really means you can test to see if one groups averages are greater than, less than, or equal to another's. From the results of those tests, which will be discussed later, the study then shifted to more data collection on real flight prices for December 18th – December 31st, 2019. From there a Two-Way ANOVA was run on these samples, which was then followed by observations and simple calculations of averages.

Results: The Two-Way ANOVA was run to see if delay averages were equal between airlines or equal between days. These results were a little unexpected. There was not enough evidence to support that average delays were different for at least one day. This means that the test concluded that there wasn't a significant difference between delay times based on which day the flights occurred. What was found though was that there was enough evidence to support that the average delay times were different for at east one airline. This image shows the results of the Two-Way ANOVA. The rows represented days, and columns represented airlines. The claim in this test was that at least one mean was different for both airlines and days. For this claim to be supported, the F value had to be higher than the F Critical value to support the claim.

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	1695.56	13	130.428	1.80111	0.07799	1.98053
Columns	843.083	3	281.028	3.88079	0.01608	2.84507
Error	2824.19	39	72.4151			

Six different Paired T-Tests were then conducted. The first three each tested Frontier versus the other airlines, with the claim that Frontier's average delays were greater than the other three airlines averages. For all three tests, there was enough evidence to say that Frontier's average delay times are statistically significantly higher than Alaska, Delta, and United Airlines average delays. Results of the test are shown in the image, where in order to support the claim the t stat had to be higher than the t critical, which it is in all three cases.

The next three Paired T-Tests were:

- Alaska vs Delta
- Alaska vs United
- Delta vs United

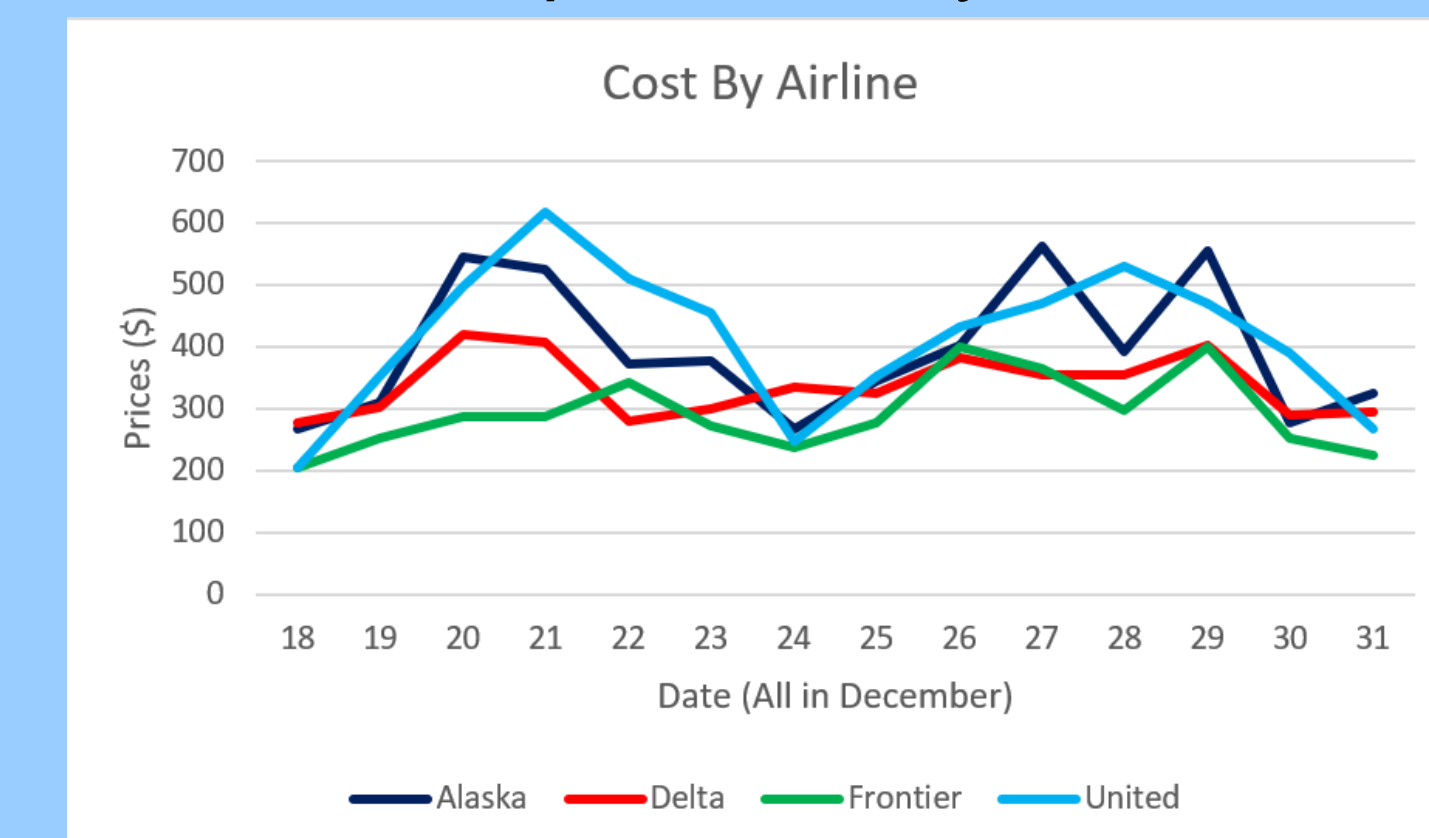
	Frontier vs Alaska	Frontier vs Delta	Frontier vs United
t Stat	1.80217	2.916267	2.04839
P(T<=t) on	0.047371	0.006013	0.030637
t Critical o	1.770933	1.770933	1.770933

These three Paired T-Tests were testing to see if there was a significant difference between the average delay times. The claim for all three tests were that the averages for both airlines being tested were equal. All three tests failed to reject the claim, this means that there isn't a statistical difference between the average delay times for Alaska, Delta, and United Airlines. By looking at this next table. It shows what all the t-tests just proved, that Frontier's averages are higher than the other three airlines, but United, Delta, and Alaska are so close that there isn't a significant statistical difference.

SUMMARY	Count	Sum	Average	Variance
Alaska	14	30.4791964	2.177085	9.919020097
Delta	14	-14.4132272	-1.02952	95.49260012
Frontier	14	132.0879188	9.434851	225.7804932
United	14	22.6577533	1.618411	16.48083123

Since delay time didn't vary by day, there was a shift in the research to consider real flight prices for each of these airlines between December18th – December 31st, 2019. Flight prices were taken from Expedia, the cheapest flight for all four airlines were chosen on each day for afternoon trips (12:00pm – 5:59pm) from San Francisco to Atlanta, and Minneapolis to San Diego.

Two-Way ANOVA was run between days and airlines for price averages and there was enough evidence to say that at least one of the days' averages were different and at least one of the airlines' averages were different. The graph below shows averages by airline per day, where United is on average the most expensive and Frontier is the cheapest. From averages calculated, the cheapest days to fly are the 18th and 24th where the most expensive days would be the 21st and the 29th.



Discussion: There were many situations where variation in the results could have come. Airports were originally selected because of the amount of traffic going through them, but smaller airports could have completely different results. Another aspect that could create some variation in the data is that airports such as Hartfield-Jackson Atlanta has Delta the only airline to fly out of concourse A while concourse D has over 10 different airlines flying into it. With only one airline flying into a concourse, it could eliminate variability of other airline delays to interfere with their arrival time. Another cause for variability in data is weather because it is so different, it could make the data more skewed. Flights are very susceptible to outside factors, which makes it difficult to create an all-around similar data pool to isolate and find which day of the week around the holidays would have the least amount of delays. This could be why the average delay between days was statistically not different.

Conclusions: From the research, it became clear that basing your decision on which day to fly based on the On-Time Arrival statistics should not be done. However, other factors such as which airline is more on time and prices would be better factors to consider. There is no statistically significant difference for average delay between days. Even though Delta tied for the most on-time airline, they have won many distinctive awards for On-Time Excellence such as the Cirium Performance Awards. Delta was also on average the second cheapest airline to fly around the holidays; the cheapest being Frontier which was also the one who averaged the most delay time. If choosing the day of flight based on the average cheapest days to fly, we would recommend to fly on December 18th or December 24th and the day to avoid would be December 21st because it is the most expensive.

Further Research: Further research that could help strengthen this study would be to explore more in depth the statistical relationship between cost of flight and date of travel. Is there statistically a closer relationship between day of flight and prices? What we would also like to investigate more is seeing if a statistical difference could be found between United, Delta, and Alaska's delay times through testing other claims.

References

(n.d.). Retrieved from <https://transtats.bts.gov/ONTIME/Arrivals.aspx>.
10th Annual Performance Awards in On-time Excellence. (n.d.). Retrieved from <https://cms.flightstats.com/media/on-time-performance-awards/#1483394269198-d937c910-654a>.
(n.d.). Retrieved from <http://atlanta-atl.worldairportguides.com/terminals.php>
Bloom, L. B. (2019, October 4). Ranked: The Best And Worst Airlines In America. Retrieved from <https://www.forbes.com/sites/laurabegleybloom/2018/03/06/ranked-the-best-and-worst-airlines-in-america/#9b452cde953b>.
Delta awarded FlightGlobal's most on-time airline as cancel-free days soar to 251. (n.d.). Retrieved from <https://news.delta.com/delta-awarded-flightglobal-s-most-time-airline-cancel-free-days-soar-251>.
Editor, C. A. (2017, December 23). Six quotes that will inspire you to travel this Christmas: eTurboNews: Trends: Travel News. Retrieved from <https://www.eturbonews.com/173704/six-quotes-will-inspire-travel-christmas/>.
Hetter, K. (2019, September 17). This is the world's busiest airport. Retrieved from <https://www.cnn.com/travel/article/worlds-busiest-airports-2018/index.html>.